RE-RUN

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.

Application Serial Number: __10/044,442__

Source: _____

Date Processed by STIC: _____

ENTERED

IFW16

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/044,442
DATE: 01/06/2005
TIME: 10:24:13

Input Set : N:\FANTU\US10044442.raw
Output Set: N:\CRF4\01052005\J044442.raw

SEQUENCE LISTING

```
 1 (1) GENERAL INFORMATION:
 2       (i) APPLICANT: Malcolm Whitman and Xin Chen
 3      (ii) TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
 4                              TGF-BETA SUPERFAMILY SIGNALLING
 5     (iii) NUMBER OF SEQUENCES: 18
 6      (iv) CORRESPONDENCE ADDRESS:
 7           (A) ADDRESSEE: Clark & Elbing LLP
 8           (B) STREET: 176 Federal Street
 9           (C) CITY: Boston
10           (D) STATE: MA
11           (E) COUNTRY: USA
12           (F) ZIP: 02110
13       (v) COMPUTER READABLE FORM:
14           (A) MEDIUM TYPE: Diskette
15           (B) COMPUTER: IBM Compatible
16           (C) OPERATING SYSTEM: DOS
17           (D) SOFTWARE: FastSEQ for Windows Version 2.0
18      (vi) CURRENT APPLICATION DATA:
```
C--> 19   **(A) APPLICATION NUMBER: US/10/044,442**
C--> 20   **(B) FILING DATE: 11-Jan-2002**
```
21           (C) CLASSIFICATION:
22     (vii) PRIOR APPLICATION DATA:
```
W--> 23   **(A) APPLICATION NUMBER: US/09/087,134**
```
24           (B) FILING DATE: 27-MAY-1998
```
W--> 25   **(A) APPLICATION NUMBER: 60/047,991**
```
26           (B) FILING DATE: 28-MAY-1997
27    (viii) ATTORNEY/AGENT INFORMATION:
28           (A) NAME: Bieker-Brady, Kristina
29           (B) REGISTRATION NUMBER: 39,109
30           (C) REFERENCE/DOCKET NUMBER: 00246/501002
31      (ix) TELECOMMUNICATION INFORMATION:
32           (A) TELEPHONE: 617-428-0200
33           (B) TELEFAX: 617-428-7045
34           (C) TELEX:
35 (2) INFORMATION FOR SEQ ID NO: 1:
36       (i) SEQUENCE CHARACTERISTICS:
37           (A) LENGTH: 1658 base pairs
38           (B) TYPE: nucleic acid
39           (C) STRANDEDNESS: single
40           (D) TOPOLOGY: linear
41      (ii) MOLECULE TYPE: cDNA
42      (ix) FEATURE:
```

```
          RAW SEQUENCE LISTING          DATE: 01/06/2005
          PATENT APPLICATION:  US/10/044,442   TIME: 10:24:13

          Input Set : N:\FANTU\US10044442.raw
          Output Set: N:\CRF4\01052005\J044442.raw
```

```
43          (A) NAME/KEY: Other
44          (B) LOCATION: 1...1
45          (D) OTHER INFORMATION: Xenopus Smad2 coding region
46      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
47   ATGTCGTCCA TCTTGCCTTT CACCCCGCCA GTAGTGAAGC GCCTGCTAGG ATGGAAGAAG    `60
48   TCTGCAAGTG GCACCACAGG AGCAGGTGGC GATGAGCAGA ACGGACAGGA AGAGAAGTGG     120
49   TGCGAAAAAG CGGTAAAGAG CTTGGTGAAA AAACTGAAGA AAACGGGACA ATTAGACGAG     180
50   CTTGAGAAGG CGATCACGAC GCAGAACTGC AACACGAAAT GCGTAACGAT ACCAAGCACT     240
51   TGCTCTGAAA TTTGGGGACT GAGTACAGCA AATACCATAG ATCAGTGGGA TACCACAGGC     300
52   CTTTACAGCT TCTCTGAACA AACCAGGTCT CTTGATGGTC GACTCCAGGT GTCTCACCGT     360
53   AAAGGATTGC CGCATGTTAT CTACTGCAGA CTGTGGCGCT GGCCAGACCT GCACAGTCAT     420
54   CATGAACTGA AAGCAATCGA AAATTGTGAA TATGCTTTTA ACCTTAAAAA AGATGAAGTT     480
55   TGTGTCAATC CATACCATTA TCAGAGGGTG GAGACACCAG TTTTACCACC TGTATTAGTT     540
56   CCACGGCACA CGGAAATCTT GACAGAGCTG CCACCTCTTG ATGACTACAC GCATTCCATT     600
57   CCAGAAAACA CTAATTTTCC TGCAGGGATT GAACCTCAGA GCAATTATAT TCCAGAAACA     660
58   CCACCTCCTG GATATATTAG TGAAGATGGA GAAACTAGCG ATCAGCAACT TAACCAAAGC     720
59   ATGGACACAG GGTCACCAGC TGAGCTGTCT CCGAGTACAC TTTCTCCAGT CAACCACAAT.    780
60   CTCGATTTGC AACCTGTCAC CTATTCGGAA CCTGCTTTTT GGTGCTCTAT AGCATACTAC     840
61   GAACTGAATC AGCGAGTAGG AGAAACTTTC CATGCATCGC AACCATCGCT TACCGTGGAC     900
62   GGCTTTACGG ACCCCTCAAA CTCTGAAAGG TTCTGCTTAG GTTTACTCTC AAATGTGAAC     960
63   CGAAATGCCA CGGTGGAAAT GACCAGGCGT CACATAGGAA GGGGTGTCCG GCTATATTAC    1020
64   ATCGGTGGAG AGGTGTTTGC AGAGTGCCTA AGTGATAGTG CTATTTTTGT TCAGAGTCCA    1080
65   AACTGTAACC AGCGATATGG ATGGCATCCA GCAACTGTAT GTAAGATTCC TCCAGGATGC    1140
66   AATCTGAAGA TTTTCAATAA TCAAGAGTTT GCGGCTCTCC TCGCTCAGTC TGTGAATCAA    1200
67   GGCTTTGAAG CAGTTTATCA GTTAACTCGA ATGTGCACCA TAAGGATGAG CTTTGTAAAA    1260
68   GGCTGGGGTG CTGAATACAG GCGACAGACC GTTACAAGCA CTCCATGCTG GATTGAGCTT    1320
69   CACCTGAATG GACCTTTGCA GTGGTTGGAC AAAGTGTTGA CACAGATGGG ATCCCCTTCA    1380
70   GTCCGCTGCT CAAGCATGTC CTAATGGTCT CCTCTTTTTA ATGTATTACC TGCGGGCGGC    1440
71   AACTGCAGTC CCAGCAACAG ACTCAATACA GCTTGTCTGT CGTAGTATTT GTGTGTGGTG    1500
72   CCCATGAACT GTTTACAATC CAAAAGAGAG AGAATAAAAA AGCAAAAACA GCACTTGAGA    1560
73   TCCCATCAAC GAAAAGCACC TTGTTGGATG ATGTTTCTGA TACTCTTAAA GTAGATCCGT    1620
74   GTATAAATGA CTCCTTACCT GGGAAAAGGG ACTTTTTC                            1658
76  (2) INFORMATION FOR SEQ ID NO: 2:
77      (i) SEQUENCE CHARACTERISTICS:
78          (A) LENGTH: 467 amino acids
79          (B) TYPE: amino acid
80          (C) STRANDEDNESS: single
81          (D) TOPOLOGY: linear
82      (ii) MOLECULE TYPE: protein
83      (ix) FEATURE:
84          (A) NAME/KEY: Other
85          (B) LOCATION: 1...1
86          (D) OTHER INFORMATION: Xenopus Smad2
87      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
88   Met Ser Ser Ile Leu Pro Phe Thr Pro Pro Val Val Lys Arg Leu Leu
89    1               5                   10                  15
90   Gly Trp Lys Lys Ser Ala Ser Gly Thr Thr Gly Ala Gly Gly Asp Glu
91                   20                  25                  30
92   Gln Asn Gly Gln Glu Glu Lys Trp Cys Glu Lys Ala Val Lys Ser Leu
```

```
93            35                      40                      45
94   Val Lys Lys Leu Lys Lys Thr Gly Gln Leu Asp Glu Leu Glu Lys Ala
95      50                      55                      60
96   Ile Thr Thr Gln Asn Cys Asn Thr Lys Cys Val Thr Ile Pro Ser Thr
97   65                      70                      75                      80
98   Cys Ser Glu Ile Trp Gly Leu Ser Thr Ala Asn Thr Ile Asp Gln Trp
99                      85                      90                      95
100  Asp Thr Thr Gly Leu Tyr Ser Phe Ser Glu Gln Thr Arg Ser Leu Asp
101                    100                     105                     110
102  Gly Arg Leu Gln Val Ser His Arg Lys Gly Leu Pro His Val Ile Tyr
103          115                     120                     125
104  Cys Arg Leu Trp Arg Trp Pro Asp Leu His Ser His His Glu Leu Lys
105      130                     135                     140
106  Ala Ile Glu Asn Cys Glu Tyr Ala Phe Asn Leu Lys Lys Asp Glu Val
107  145                     150                     155                     160
108  Cys Val Asn Pro Tyr His Tyr Gln Arg Val Glu Thr Pro Val Leu Pro
109                    165                     170                     175
110  Pro Val Leu Val Pro Arg His Thr Glu Ile Leu Thr Glu Leu Pro Pro
111                    180                     185                     190
112  Leu Asp Asp Tyr Thr His Ser Ile Pro Glu Asn Thr Asn Phe Pro Ala
113          195                     200                     205
114  Gly Ile Glu Pro Gln Ser Asn Tyr Ile Pro Glu Thr Pro Pro Pro Gly
115      210                     215                     220
116  Tyr Ile Ser Glu Asp Gly Glu Thr Ser Asp Gln Gln Leu Asn Gln Ser
117  225                     230                     235                     240
118  Met Asp Thr Gly Ser Pro Ala Glu Leu Ser Pro Ser Thr Leu Ser Pro
119                    245                     250                     255
120  Val Asn His Asn Leu Asp Leu Gln Pro Val Thr Tyr Ser Glu Pro Ala
121                    260                     265                     270
122  Phe Trp Cys Ser Ile Ala Tyr Tyr Glu Leu Asn Gln Arg Val Gly Glu
123          275                     280                     285
124  Thr Phe His Ala Ser Gln Pro Ser Leu Thr Val Asp Gly Phe Thr Asp
125      290                     295                     300
126  Pro Ser Asn Ser Glu Arg Phe Cys Leu Gly Leu Leu Ser Asn Val Asn
127  305                     310                     315                     320
128  Arg Asn Ala Thr Val Glu Met Thr Arg Arg His Ile Gly Arg Gly Val
129                    325                     330                     335
130  Arg Leu Tyr Tyr Ile Gly Gly Glu Val Phe Ala Glu Cys Leu Ser Asp
131                    340                     345                     350
132  Ser Ala Ile Phe Val Gln Ser Pro Asn Cys Asn Gln Arg Tyr Gly Trp
133          355                     360                     365
134  His Pro Ala Thr Val Cys Lys Ile Pro Pro Gly Cys Asn Leu Lys Ile
135      370                     375                     380
136  Phe Asn Asn Gln Glu Phe Ala Ala Leu Leu Ala Gln Ser Val Asn Gln
137  385                     390                     395                     400
138  Gly Phe Glu Ala Val Tyr Gln Leu Thr Arg Met Cys Thr Ile Arg Met
139                    405                     410                     415
140  Ser Phe Val Lys Gly Trp Gly Ala Glu Tyr Arg Arg Gln Thr Val Thr
141                    420                     425                     430
```

```
           RAW SEQUENCE LISTING              DATE: 01/06/2005
           PATENT APPLICATION:  US/10/044,442     TIME: 10:24:13

           Input Set : N:\FANTU\US10044442.raw
           Output Set: N:\CRF4\01052005\J044442.raw

142  Ser Thr Pro Cys Trp Ile Glu Leu His Leu Asn Gly Pro Leu Gln Trp
143          435             440             445
144  Leu Asp Lys Val Leu Thr Gln Met Gly Ser Pro Ser Val Arg Cys Ser
145      450             455             460
146  Ser Met Ser
147  465
149  (2) INFORMATION FOR SEQ ID NO: 3:
150      (i) SEQUENCE CHARACTERISTICS:
151          (A) LENGTH: 194 amino acids
152          (B) TYPE: amino acid
153          (C) STRANDEDNESS: single
154          (D) TOPOLOGY: linear
155      (ii) MOLECULE TYPE: protein
156      (ix) FEATURE:
157          (A) NAME/KEY: Other
158          (B) LOCATION: 1...1
159          (D) OTHER INFORMATION: Xenopus Smad2 MH2 domain
160      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
161  Trp Cys Ser Ile Ala Tyr Tyr Glu Leu Asn Gln Arg Val Gly Glu Thr
162   1           5               10              15
163  Phe His Ala Ser Gln Pro Ser Leu Thr Val Asp Gly Phe Thr Asp Pro
164          20              25              30
165  Ser Asn Ser Glu Arg Phe Cys Leu Gly Leu Leu Ser Asn Val Asn Arg
166          35              40              45
167  Asn Ala Thr Val Glu Met Thr Arg Arg His Ile Gly Arg Gly Val Arg
168      50              55              60
169  Leu Tyr Tyr Ile Gly Gly Glu Val Phe Ala Glu Cys Leu Ser Asp Ser
170  65              70              75              80
171  Ala Ile Phe Val Gln Ser Pro Asn Cys Asn Gln Arg Tyr Gly Trp His
172              85              90              95
173  Pro Ala Thr Val Cys Lys Ile Pro Pro Gly Cys Asn Leu Lys Ile Phe
174              100             105             110
175  Asn Asn Gln Glu Phe Ala Ala Leu Leu Ala Gln Ser Val Asn Gln Gly
176          115             120             125
177  Phe Glu Ala Val Tyr Gln Leu Thr Arg Met Cys Thr Ile Arg Met Ser
178      130             135             140
179  Phe Val Lys Gly Trp Gly Ala Glu Tyr Arg Arg Gln Thr Val Thr Ser
180  145             150             155             160
181  Thr Pro Cys Trp Ile Glu Leu His Leu Asn Gly Pro Leu Gln Trp Leu
182              165             170             175
183  Asp Lys Val Leu Thr Gln Met Gly Ser Pro Ser Val Arg Cys Ser Ser
184          180             185             190
185  Met Ser
187  (2) INFORMATION FOR SEQ ID NO: 4:
188      (i) SEQUENCE CHARACTERISTICS:
189          (A) LENGTH: 1401 base pairs
190          (B) TYPE: nucleic acid
191          (C) STRANDEDNESS: single
192          (D) TOPOLOGY: linear
```

```
193      (ii)  MOLECULE TYPE: cDNA
194      (ix)  FEATURE:
195            (A)  NAME/KEY: Other
196            (B)  LOCATION: 1...1
197            (D)  OTHER INFORMATION: Human Smad2 coding region
198      (xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 4:
199  ATGTCGTCCA TCTTGCCATT CACGCCGCCA GTTGTGAAGA GACTGCTGGG ATGGAAGAAG      60
200  TCAGCTGGTG GGTCTGGAGG AGCAGGCGGA GGAGAGCAGA ATGGGCAGGA AGAAAAGTGG     120
201  TGTGAGAAAG CAGTGAAAAG TCTGGTGAAG AAGCTAAAGA AAACAGGACG ATTAGATGAG     180
202  CTTGAGAAAG CCATCACCAC TCAAAACTGT AATACTAAAT GTGTTACCAT ACCAAGCACT     240
203  TGCTCTGAAA TTTGGGGACT GAGTACACCA AATACGATAG ATCAGTGGGA TACAACAGGC     300
204  CTTTACAGCT TCTCTGAACA AACCAGGTCT CTTGATGGTC GTCTCCAGGT ATCCCATCGA     360
205  AAAGGATTGC CACATGTTAT ATATTGCCGA TTATGGCGCT GGCCTGATCT TCACAGTCAT     420
206  CATGAACTCA AGGCAATTGA AAACTGCGAA TATGCTTTTA ATCTTAAAAA GGATGAAGTA     480
207  TGTGTAAACC CTTACCACTA TCAGAGAGTT GAGACACCAG TTTTGCCTCC AGTATTAGTG     540
208  CCCCGACACA CCGAGATCCT AACAGAACTT CCGCCTCTGG ATGACTATAC TCACTCCATT     600
209  CCAGAAAACA CTAACTTCCC AGCAGGAATT GAGCCACAGA GTAATTATAT TCCAGAAACG     660
210  CCACCTCCTG GATATATCAG TGAAGATGGA GAAACAAGTG ACCAACAGTT GAATCAAAGT     720
211  ATGGACACAG GCTCTCCAGC AGAACTATCT CCTACTACTC TTTCCCCTGT TAATCATAGC     780
212  TTGGATTTAC AGCCAGTTAC TTACTCAGAA CCTGCATTTT GGTGTTCAAT AGCATATTAT     840
213  GAATTAAATC AGAGGGTTGG AGAAACCTTC CATGCATCAC AGCCCTCACT CACTGTAGAT     900
214  GGCTTTACAG ACCCATCAAA TTCAGAGAGG TTCTGCTTAG GTTTACTCTC CAATGTTAAC     960
215  CGAAATGCCA CGGTAGAAAT GACAAGAAGG CATATAGGAA GAGGAGTGCG CTTATACTAC    1020
216  ATAGGTGGGG AAGTTTTTGC TGAGTGCCTA AGTGATAGTG CAATCTTTGT GCAGAGCCCC    1080
217  AATTGTAATC AGAGATATGG CTGGCACCCT GCAACAGTGT GTAAAATTCC ACCAGGCTGT    1140
218  AATCTGAAGA TCTTCAACAA CCAGGAATTT GCTGCTCTTC TGGCTCAGTC TGTTAATCAG    1200
219  GGTTTTGAAG CCGTCTATCA GCTAACTAGA ATGTGCACCA TAAGAATGAG TTTTGTGAAA    1260
220  GGGTGGGGAG CAGAATACCG AAGGCAGACG GTAACAAGTA CTCCTTGCTG GATTGAACTT    1320
221  CATCTGAATG GACCTCTACA GTGGTTGGAC AAAGTATTAA CTCAGATGGG ATCCCCTTCA    1380
222  GTGCGTTGCT CAAGCATGTC A                                               1401
224  (2) INFORMATION FOR SEQ ID NO: 5:
225      (i)  SEQUENCE CHARACTERISTICS:
226           (A)  LENGTH: 467 amino acids
227           (B)  TYPE: amino acid
228           (C)  STRANDEDNESS: single
229           (D)  TOPOLOGY: linear
230      (ii)  MOLECULE TYPE: protein
231      (ix)  FEATURE:
232           (A)  NAME/KEY: Other
233           (B)  LOCATION: 1...1
234           (D)  OTHER INFORMATION: Human Smad2
235      (xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 5:
236  Met Ser Ser Ile Leu Pro Phe Thr Pro Pro Val Val Lys Arg Leu Leu
237   1               5                  10                      15
238  Gly Trp Lys Lys Ser Ala Gly Gly Ser Gly Gly Ala Gly Gly Gly Glu
239               20                  25                      30
240  Gln Asn Gly Gln Glu Glu Lys Trp Cys Glu Lys Ala Val Lys Ser Leu
241           35                  40                      45
242  Val Lys Lys Leu Lys Lys Thr Gly Arg Leu Asp Glu Leu Glu Lys Ala
```

VERIFICATION SUMMARY
PATENT APPLICATION: **US/10/044,442**

DATE: 01/06/2005
TIME: 10:24:14

Input Set : **N:\FANTU\US10044442.raw**
Output Set: **N:\CRF4\01052005\J044442.raw**

L:19 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:20 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:25 M:238 W: Alpha Fields not Ordered, Reordered [(A) APPLICATION NUMBER:] of (1)(vii)